The Computerworld Honors Program

CASE STUDY

ORGANIZATION:

George Washington University

PROJECT NAME:

Tiered Storage and Information Lifecycle Management

LOCATION:
Washington, DC, United States

YEAR:
2006

STATUS:
Laureate

CATEGORY:
Education and Academia

NOMINATING COMPANY:
EMC

Summary

For The George Washington University (GW), the events of 9/11 were an especially loud wake-up call. With a main campus and datacenter located in Foggy Bottom (Washington, D.C.) just four blocks from the White House, the University's Information Systems and Services (ISS) department recognized that its existing disaster recovery approach would not provide adequate protection in case of a major disruption to the primary datacenter. Plus, with GW's fast-growing IT requirements, the datacenter was "bursting at the seams."

To address these issues, GW dramatically expanded its main datacenter infrastructure, centralized previously disparate pockets of data storage across its three campuses and, most important, established advanced business continuity. Following the project tagline "Choosing Continuity, Not Recovery," the ISS staff led by David Swartz, GW's Chief Information Officer (CIO), implemented a dual datacenter strategy that, in case of a major disruption at either site, ensures restoration of critical data services in minutes with zero data loss.

Introductory Overview

With a main campus in downtown Washington, D.C., GW is in a uniquely advantageous as well as vulnerable location. Located in Foggy Bottom, just four blocks from the White House and across the street from the U.S. Department of State, it lies within a critically important area. That's why, after 9/11, GW's ISS staff took a hard look at its disaster recovery strategy and, at the same time, its enterprise-wide data management and storage infrastructure. They immediately saw changes that needed to be made.

First and foremost, GW needed to radically transform its disaster recovery strategy. All critical enterprise data was stored and backed up at the primary datacenter downtown. The only redundancy was daily deliveries of tape backups to a remote offsite location. If the primary datacenter was disrupted or damaged, IT operations would be halted for weeks, dramatically affecting the ability of the University to fully function.

The ISS team found other critical IT needs as well. Demands for GW's computing resources grew dramatically throughout the late 1990s and datacenter capacity had reached its limits.

THE COMPUTERWORLD HONORS PROGRAM

CASE STUDY

ORGANIZATION:
George Washington University

PROJECT NAME:
Tiered Storage and Information Lifecycle Management

LOCATION:
Washington, DC, United States

YEAR:
2006

STATUS:
Laureate

CATEGORY:
Education and Academia

NOMINATING COMPANY:
EMC

Ron Bonig, GW's Deputy CIO, said, "While our user base and number of supported applications were expanding, the datacenter walls were not."

The University needed not only increased, "industrial-strength" server resources, but it needed to completely rethink, redesign and expand its data storage infrastructure. GW had what Bonig termed a "balkanized" model of data storage, referencing the historic lack of coordination among the Balkan countries of southeastern Europe. Beyond the datacenter, each academic and administrative department maintained separate "islands of information," with no centralized data management, storage or backup. Even in the datacenter, each legacy server had its own, direct-attached storage.

To address these issues, GW's Information Systems and Services department launched a massive program to transform GW computing resources and embody the new mantra of "Choosing Continuity, Not Recovery." This included:

·Centralizing and expanding both enterprise and departmental data storage,

·Implementing information lifecycle management (ILM) so that different categories of data – from mission-critical enterprise applications to long-term record archives – could reside on the most appropriate and cost-effective tier of storage while remaining online and quickly accessible,

·Establishing 24x7 business continuity with dual datacenters for continuous, remote replication of key applications, uninterrupted IT services, rapid recovery and near-zero data loss, should either datacenter experience an outage.

Today, this vision has been realized. Even in a worst-case scenario, redundant IT operations would continue, with full recovery of critical applications in hours rather than days. GW's IT infrastructure is able to easily and cost-effectively scale to support an ever-growing number of financial, administrative and student-focused information services. It has virtually eliminated the risk of interruption, meeting the "five nines" standard for network availability – 99.999 percent uptime.

Benefits

GW's transformation of its data storage infrastructure and implementation of business continuity has benefited every area and aspect of university life. The change is most dramatic in the area that spurred the ISS division to action – business continuity. Previously, GW had only one production datacenter, with all critical enterprise applications stored and backed up on-site. The only redundancy was daily delivery of backup tapes to a remote cold-site location belonging to a contracted vault vendor.

Aside from the cost, time and inconvenience involved with transporting backup tapes in this manner, GW's previous disaster recovery strategy left the vulnerable datacenter in an untenable position. If it were completely destroyed by a catastrophic event, the University would literally have had to build a complete new datacenter and catch up on certain tasks that would need to be temporarily halted during the months of rebuilding. Worse yet, GW would essentially be shut down for that period of time. Even with the datacenter intact, it would have taken five to seven days to fully restore data and services.

With GW's new business continuity infrastructure, that scenario has been eliminated. The ISS group now has two, fully mirrored datacenters: the downtown site and a second site 35 miles

ORGANIZATION:
George Washington University

PROJECT NAME:
Tiered Storage and Information Lifecycle Management

LOCATION:
Washington, DC, United States

YEAR:
2006

STATUS:
Laureate

CATEGORY:
Education and Academia

NOMINATING COMPANY:
EMC

away in northern Virginia. Using EMC Symmetrix Remote Data Facility (SRDF) software, data is continuously replicated between the two sites. With complete hardware, software and communications redundancy, if either site is knocked out, there would be minimal impact on "customers" – GW students, faculty and administrative staff. Because the University's recovery point objective (RPO) has improved, there would be near-zero data loss, and the most mission-critical data and applications would be restored in less than two hours and in some cases under an hour. All of the University's Oracle-based financial applications would be up and running at the both sites within eight hours, and all ISS services would be back to normal within 24 hours.

Adding further protection, the ISS team connected the two datacenters via a 100-mile long redundant optical fiber loop. It leaves the Foggy Bottom datacenter from two different points and takes two, widely divergent routes to the Virginia datacenter. That way, if one path fails, there is the assurance of having a second active connection. Among other potential problems, it overcomes what Ron Bonig calls the problem of the "cable-seeking backhoe" – the kind of random, cable company construction and repair accidents that can cut communications links unless a separate, redundant path is available.

Having two redundant datacenters has also greatly enhanced the ISS group's ability to do testing and development. Previously, with outsourced disaster recovery, staff members would have to drive to the disaster recovery provider's datacenter 100 miles away to perform testing. Plus, they were limited to being able to carry out testing only a couple of times each year, and those testing events took a great deal of planning time. Now, the ISS team handles testing and development at their secondary site, any time they want and without out any impact on the production environment. As a result, instead of testing server failover and backup systems once or at most four times per year, they test them every time they make a change in the datacenter infrastructure. This more frequent testing schedule improves the overall reliability of GW's IT infrastructure and allows the University to further extend the value of its business continuity investments.

In conjunction with implementing business continuity was the centralization of enterprise and departmental data storage at the ISS datacenters. The ISS team eliminated many of the University's disparate "islands of information" across the campuses, bringing them, in a sense, all under one roof. For the first time, the ISS group has a 360-degree view of GW's total information resources. The benefit for the ISS group is that it can effectively manage these resources. For the GW community, this means that departmental technologists can focus on technologies peculiar to their specific line of business rather than managing IT infrastructure. For example, before centralizing infrastructure management with ISS, the University's Associate Librarian for technology spent 80 to 90 percent of his staff's time performing necessary server and storage maintenance. Since the transition, library technologists spend the majority of their time delivering library specific content to the library constituents.

With the centralization of data and application came the challenge – and opportunity – of implementing ILM. Where once there was only one level of top-tier storage, ISS creating three tiers of storage, enabling the staff to stratify data and applications by levels of criticality and accessibility. They consist of:

•Top-Tier (Tier 1) critical enterprise systems, including the extensive SCT Banner suite of University administration and student record applications, Oracle-based financial systems, and GW's SUN One email, are stored on EMC Symmetrix DMX storage, with EMC SRDF software providing synchronous and continuous replication between the dual datacenters. EMC

THE COMPUTERWORLD HONORS PROGRAM

CASE STUDY

ORGANIZATION:
George Washington University

PROJECT NAME:
Tiered Storage and Information Lifecycle Management

LOCATION:
Washington, DC, United States

YEAR:
2006

STATUS:
Laureate

CATEGORY:
Education and Academia

NOMINATING COMPANY:
EMC

TimeFinder software is used to create data replica used for remote backups, patch testing, and upgrade tests.

•EMC CLARiiON Fiber Channel and ATA technologies are used for GW's Tier 2 and Tier 3 storage, respectively. This includes storage for non-enterprise, departmental applications, academic computing, and departmental file-sharing. EMC MirrorView software is applied to these tiers of storage similar to the way EMC SRDF is applied to GW's Tier 1 storage to provide synchronous mirroring between datacenters. While the performance needs of this data are not quite as high as Tier 1, high-availability and disaster recovery is still a top requirement. EMC SnapView software is used to create full copies and point-in-time images of data for remote backups, patch testing and upgrade tests.

•Tier 4 CAS (Content Addressed Storage) provided by EMC Centera technology is the chosen architecture for GW's document archiving. Images of contracts, invoices, and financial documents the University doesn't need to access on a regular basis but still must have available, protected, and preserved in their original format (protected against false modification) are stored here. These documents were previously stored in hard copy, accessible by very few people, expensive to store, and subject to loss and physical deterioration. Now, they are remotely accessible, secured, and protected from physical damage via local and remote Centera replication.

As a university that follows a business model of management and operations, GW will benefit financially from ILM. Only the most important data and applications need the high-response services (accessibility, backup and restoration windows) that come at a higher cost. ILM provides a way to peg storage costs to the mission-criticality of data, thereby saving the University carefully husbanded resources. Also, in keeping with its business model, GW's Tier 4 CAS archival storage complies with Sarbanes-Oxley regulations for long-term storage and access to financial records.

With expanded, centralized IT resources, GW has ensured a solid foundation for continuing to meet the growing computing needs of the university's community. For example, GW has deployed the GWorld student credit and security card. Students need their GWorld cards to get into many dorms, libraries and other university facilities, but they can also use them to buy books, meals, merchandise at on- and off-campus shops, and tickets to concerts or sports events. Other new GW services include the Web-based GWired online network and MyGW Portal, giving students links to University news and resources.

The Importance of Technology

GW's ILM strategy involves four levels of data storage – each of which utilizes a different approach to business continuity. The most critical information is replicated on a continuous basis using EMC SRDF/Synchronous (SRDF/S) mirroring software between the two datacenters. Further, no transaction, e.g., recording of a student registration, is completed until it is written to both sites. Continuous replication ensures that, if one datacenter goes down, no data is lost and data services from the affected site can be restored in minutes.

To add additional levels of protection and functionality, Tier 1 data is copied nightly to business continuity volumes (BCVs) -- full point-in-time copies of data -- using EMC TimeFinder software. These BCV volumes are mounted to a backup host, backed up to tape, duplicated, then encrypted and stored off-site. This use of encryption ensures that, in the unlikely event that the

ORGANIZATION:
George Washington University

PROJECT NAME:
Tiered Storage and Information Lifecycle Management

LOCATION:
Washington, DC, United States

YEAR:
2006

STATUS:
Laureate

CATEGORY:
Education and Academia

NOMINATING COMPANY:
EMC

tapes are intercepted, they could not be read. The information would still be "untouchable."

GW maintains two EMC CLARiiON CX700 storage systems for mid-tier applications. With EMC CLARiiON technology, GW is able to incorporate a mixture of disk technology in order to make available different levels of performance for different application needs. Currently, GW offers a mixture of high-speed Fibre Channel, SATA, and PATA disk drives in both of its CLARiiON systems. The CLARiiONs support an online Logical Unit Number (LUN) migration between tiers. So, if a system was placed on the PATA drives, and it is later discovered that the application needs Fibre Channel disk speeds, the application can be migrated to this faster disk without suffering downtime.

For mid-tier applications deemed "remote DR necessary," EMC MirrorView/Synchronous (MirrorView/S) provides synchronous remote data replication for applications on the CLARiiON. For high-availability applications, EMC SnapView clones and/or snapshots are used in ways similar to BCV volumes on the Symmetrix DMX.

GW uses EMC Centera content addressed storage, which is specifically designed for archived, fixed content that must not be altered. While still accessible in seconds, GW's fixed content is fully protected and is stored on a lower-cost medium compared with mid-tier and high-end storage. To further protect fixed content, GW relies on Centera replication software to mirror Centera content between the two datacenters. In addition, the ISS team utilizes EMC Documentum software to manage records stored on Centera in compliance with Sarbanes-Oxley regulations. The GW Comptroller's office uses EMC DiskXtender software to archive long retention documents, such as tax paperwork, payroll information, and endowment documentation.

Beyond data storage hardware and software functionality, communications technology was critical to GW's business continuity strategy. To achieve the necessary bandwidth, they connected the two datacenters via two separate, redundant fiber optic cable paths incorporating Dense Wavelength Division Multiplexing (DWDM) technology. DWDM enabled the ISS group to simplify implementation of its storage area network (SAN) and establish redundant links from the SAN to all storage systems, hosts and tape devices.

Originality

Within two years, GW has gone from an almost non-existent disaster recovery posture to full business continuity. Now, should a catastrophic event strike either the central or "hot" remote datacenter, computing operations would be minimally affected. The university's most critical applications can be restored in less than two hours with near-zero data loss – a far cry from having all IT systems down for days or weeks, with University operations drawn to a halt in the meantime.

At the same time, GW changed a decentralized storage environment largely inaccessible to much of the University community to a centrally managed and tiered IT infrastructure that is widely depended on by students, faculty and administrators.

Either part of GW's IT transformation – the centralization and tiering of data storage or the implementation of business continuity – would have been a major undertaking. Carrying out both projects as part of one coordinated effort is a truly exceptional achievement.

THE COMPUTERWORLD HONORS PROGRAM

CASE STUDY

ORGANIZATION:
George Washington University

PROJECT NAME:
Tiered Storage and Information Lifecycle Management

LOCATION:
Washington, DC, United States

YEAR:
2006

STATUS:
Laureate

CATEGORY:
Education and Academia

NOMINATING COMPANY:
EMC

The most innovative and unusual aspect of GW's business continuity is the implementation of the dual datacenters and the many "failsafe" features that went into their design. The ISS team created almost identical datacenters, each with essentially the same Sun Solaris server farm and tiered data storage and backup systems. In linking the two sites, the team leased 100 miles of "dark" fiber in two segments. Each leaves the downtown datacenter from different points, crosses different sections of Washington, goes over two different bridges and routes up to 35 miles apart, to reach the northern Virginia datacenter target site, again at two different points. This mimics the technique used by personal security organizations to conceal the exact route of an important official, using multiple vehicles and routes to keep potential attackers off-guard.

This storage hardware redundancy is supported by data replication within milliseconds of data transactions between GW's top-tier storage systems at the two datacenters, using EMC SRDF and MirrorView data replication software products. Thus, financial or other transactions are not considered complete until they are written to systems at both datacenters. This added level of redundancy means that, if either datacenter goes down, the other continues to function without interruption.

Success

Happily, since implementing business continuity, GW has not experienced the kind of catastrophic event its infrastructure is ready to address. However, the entire project has been a major success in serving all University constituencies. Than includes students, administrators, faculty, and the ISS department itself.

Most important, GW now has the assurance of continuous access to data and applications. In the event of a serious disruption to one of its datacenters, it would be only a matter of hours, not days, until its services are fully restored and in operation. At the same time, the project has dramatically improved ISS storage capacity, data management capabilities, efficiency, and performance.

This success can be measured in such operations benefits as the major reduction in data backup times, reduced maintenance windows and increased support for systems development. For example, a full remote backup of its Oracle financial applications has been cut from eight hours to less than 10 minutes.

With the consolidation of enterprise and departmental data storage, the ISS team successfully eliminated the University's disparate "islands of information" across the campuses. For the first time, the ISS group has a 360-degree view of GW's total information resources. The benefit for the ISS group is that they can effectively and centrally manage all of these resources. For the GW community, it means that people can access valuable repositories of information that they never knew existed.

In addition, the performance of GW systems has reached new heights. The University's network has achieved "five nines" performance, that is, 99.999% uptime. Its email service, for example, has experienced only two hours of downtime in the past 15 months. And the dark fiber network between the Washington, D.C. and northern Virginia datacenters has never had a full outage being able to operate on one link at a time in case of a fiber cut.

Another key measure of the project's success has been in return on investment (ROI). The annual operating costs for the new infrastructure are approximately the same as the old practice of

ORGANIZATION:
George Washington University

PROJECT NAME:
Tiered Storage and Information Lifecycle Management

LOCATION:
Washington, DC, United States

YEAR:
2006

STATUS:
Laureate

CATEGORY:
Education and Academia

NOMINATING COMPANY:
EMC

outsourcing disaster recovery. However, it provides a vastly improved level of protection against loss of data and dramatically increased capacity and availability of applications and ISS services.

Difficulty

The challenges in implementing business continuity were, considering the size and scope of the project, managed with ease by the ISS team. As with any multi-million dollar project, GW's Board of Trustees needed to approve the project and expenditures. Especially in the post-9/11 environment, the urgency and necessity of protecting the University's vital data resources and operations were clear and compelling. In addition, the ISS team was able to show that, by simple reallocating funds previously paid out to the third-party disaster recovery provider, GW would receive enormous benefits in terms of improved data management, access and capacity.

In reducing the overall costs of implementing the second datacenter – the largest single expense in the budget -- Ron Bonig's organization received an unexpected windfall. Just when it became clear that they needed more space, a telecommunications company directly across the street from GW's northern Virginia campus vacated their headquarters building, including a 10,000-square-foot, semi-prepped data test center. Although not a full production datacenter, it had the space and many of the amenities, including air conditioning and raised floor, which were needed for the target datacenter. GW bought the building, and the ISS group got the funds for the necessary upgrades, including the installation of dark fiber links to the primary site. Bonig estimates this saved the team one million dollars in project costs.

Once it got the funding for the project, the team met with some technical issues around implementing the remote server clusters at the remote site. However, these were resolved with the consulting assistance of GW's three primary vendors involved in the project, EMC, Cisco and Sun.

While getting the necessary licensing and approvals took time, these were accomplished relatively quickly. Ron Bonig noted that, at one faculty meeting, he was asked when the cutover was going to take place. He was happy to point out that it had already happened, proving that the actual launch of the new infrastructure was transparent to the non-IT University community.